Published online 28 April 2008 *Nucleic Acids Research, 2008, Vol. 36, Web Server issue* ***W233–W238***
doi:10.1093/nar/gkn216

The RosettaDock server for local protein–protein docking

Sergey Lyskov[1] and Jeffrey J. Gray[1,2,*]

[1]Department of Chemical and Biomolecular Engineering and [2]Program in Molecular and Computational Biophysics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA

Received January 31, 2008; Revised March 25, 2008; Accepted April 9, 2008

ABSTRACT

The RosettaDock server (http://rosettadock.graylab.jhu.edu) identifies low-energy conformations of a protein–protein interaction near a given starting configuration by optimizing rigid-body orientation and side-chain conformations. The server requires two protein structures as inputs and a starting location for the search. RosettaDock generates 1000 independent structures, and the server returns pictures, coordinate files and detailed scoring information for the 10 top-scoring models. A plot of the total energy of each of the 1000 models created shows the presence or absence of an energetic binding funnel. RosettaDock has been validated on the docking benchmark set and through the Critical Assessment of PRedicted Interactions blind prediction challenge.

INTRODUCTION

Protein–protein interactions underlie many basic biological processes, from signaling and regulation to recognition. Protein–protein docking, the task of predicting the 3D structure of a protein–protein complex from its component structures, is useful in the absence of an experimental structure to provide insights into the molecular function of proteins such as the basis for recognition, affinity and specificity (1).

Several protein–protein docking servers are available on the Internet, including ClusPro (2), GRAMM-X (3) and ZDOCK (4) based on fast-Fourier transform methods for grid matching; PatchDock and SymmDock (5) based on shape complementarity principles and symmetry restrictions; and Hex based on spherical harmonic representations (6). These servers are fast and allow global docking searches; however, the atomic-level accuracy of the models are limited by the course-grained representation of the proteins.

RosettaDock is a structure-prediction-based program, which searches the rigid-body and side-chain conformational space of two interacting proteins to find a minimum free-energy complex structure (7). RosettaDock has been highly successful in the blind prediction challenge of the Critical Assessment of PRedicted Interactions (CAPRI) (8), producing several structures that were the most atomically accurate models submitted by any group in the CAPRI challenge. Two limitations of RosettaDock have been that (i) the command-line interface can be difficult to use and (ii) it requires significant computational time to generate all-atom models, typically requiring a cluster of computers.

To make the computation available to a broader community, we have developed a RosettaDock server (http://rosettadock.graylab.jhu.edu), where the interface is simple and the computing resources are provided. Currently, the computing cost requires us to limit the public use to local searches near user-provided starting conformations [$\sim$30 Å root mean-squared deviation (r.m.s.d). of C_α atoms]. Local searches are useful for refining top-ranked models from global searches by other docking methods or for searching for conformations given constraints provided by experimental data such as site-directed mutagenesis effects on binding affinity.

PROCESSING METHOD

RosettaDock is a multi-start, multi-scale Monte Carlo-based algorithm, which has been described previously (7). The low-resolution phase of the search includes cycles of random rigid-body perturbations with a course-grained representation of side chains as single pseudo-atoms. The high-resolution (all-atom, including hydrogens) phase of the search includes smaller rigid-body perturbations, side-chain optimization via rotamer packing and continuous minimization (9), and explicit gradient-based minimization of the rigid-body displacement. Scoring in the low-resolution phase includes residue–residue contacts and bumps, knowledge-based terms for residue environment and residue–residue pair propensities (7) and for antibody–antigen targets, a score to favor interactions with antibody complementarity determining regions (10). In the high-resolution phase, the energy is dominated by van

*To whom correspondence should be addressed. Tel: +1 410 516 5313; Fax: +1 410 516 5510; Email: jgray@jhu.edu

der Waals energies (7), orientation-dependent hydrogen bonding (11), implicit Gaussian solvation (12), side-chain rotamer probabilities (13) and a low-weighted electrostatics energy (7). For a local docking perturbation run performed by the server, 1000 independent simulations are carried out to generate an ensemble of models.

INPUTS AND OUTPUTS

Input

Structures of the docking partners are uploaded in the standard Protein Data Bank (PDB) (14) coordinate file format as two separate files or as a single file with the docking partners separated by a TER record. Since the RosettaDock server performs a local docking search near the given starting conformation, the uploaded coordinate files must provide a reasonable estimate for the starting position. The protein partners should be placed near contact (but not overlapping) with the relevant patches of the proteins facing each other.

Several initial checks are performed on the uploaded coordinate files, including checking the distance between docking partners, the total number of residues and the complete presence of all backbone atoms. The initial distance between C_α atoms in different docking partners should not be less than 5 Å to avoid initial collisions, which can cause numerical instabilities. The total number of residues in the proteins should be between 8 and 600. Proteins with less than eight residues (or even more) are unlikely to produce meaningful results, since backbone flexibility of short peptides is important but not captured by the algorithm. Protein pairs over 600 residues are prohibitively computationally expensive due to the all-atom energy calculations performed; such proteins should be manually truncated to isolate the putative interacting (sub-)domain. Coordinate files with multiple structural models (e.g. alternate NMR solutions) are not allowed. If the uploaded files fail any of these criteria, the user is notified immediately with an appropriate error message.

The user can optionally specify protein names and an email address for notification when the docking task has finished.

Output

Figure 1 shows a representative output page from the RosettaDock server. The server outputs the 10 best-scoring structures with pictures and coordinate files in rank order by energy. Each model output file includes the scoring data of individual energy terms [van der Waals, solvation, hydrogen bonding energies, etc.; see ref. (15) for notation] for the whole-protein complex as well as residue-by-residue breakdowns and intermolecular residue-pair contributions. In addition, the server returns a plot of the energies of 1000 structures created during the docking run versus the r.m.s.d. from the starting input conformation. The presence or absence of a 'docking funnel', where many low-scoring decoys have similar r.m.s.d. values indicating similar conformations, can inform the user of the convergence of the run and by extension the confidence in the provided solutions (16). Finally, raw scoring data for the 1000 decoys is provided as a flat text file. For deep analysis, the full set of 1000 decoys is provided as compressed archive files. Scientists testing their own scoring or refinement procedures may use these structures as starting configurations. Finally, a link is provided to the documentation page, which explains the output in detail, including a breakdown of the scoring terms found in the coordinate files.

SYSTEM ARCHITECTURE

Since a docking computation can require days even on multiprocessor clusters, the practical implementation is to separate the front-end web process from the computation daemon and engine. Figure 2 shows the implementation of the server architecture. The front-end web server, implemented in Python using TurboGears (http://turbogears.org), provides results upon request for users and enters docking tasks into a MySQL database once the submitted input files pass initial checks. A back-end daemon pulls tasks from the queue in the MySQL database, translates the docking task into a Rosetta++ command-line [including detecting antibody sequences to activate antibody options (10)] and submits a job to a Condor (http://www.cs.wisc.edu/condor) queue. The Condor system runs the job as time is available on a 200-processor Linux cluster, which is shared with ongoing research tasks from our lab (typically only a fraction of the cluster is used by the server). Finally, the back-end daemon periodically detects the status of the job to report, and eventually enters the complete set of results into the MySQL database.

The server is designed to be able to utilize diverse sources of computational power. The Condor queue is extendable to heterogeneous pools of asynchronous computers, and the submission task could even be switched to distribute computing platforms such as BOINC (http://boinc.berkeley.edu). In this way, we hope to eventually be able to provide adequate computing power for a large user base or to be able to distribute the server code to high-demand users who might want to run jobs on their own in-house facilities.

SERVER PERFORMANCE

Since the RosettaDock web server opened in April 2007, over 150 individuals have used the web server for more than 800 docking jobs. Jobs typically require about 65 processor-hours and results are typically complete within a few days of submission, although the time will vary with the protein sizes, the server queue and the lab's current cluster load. Users are restricted to five jobs in the queue to speed access for all users. The website is free and open to all users with no login requirement.

Accuracy of the RosettaDock server

In a large-scale test of RosettaDock, the program was used to re-dock protein–protein complexes using either bound or unbound components (7). When locally docking unbound components in a location near the native

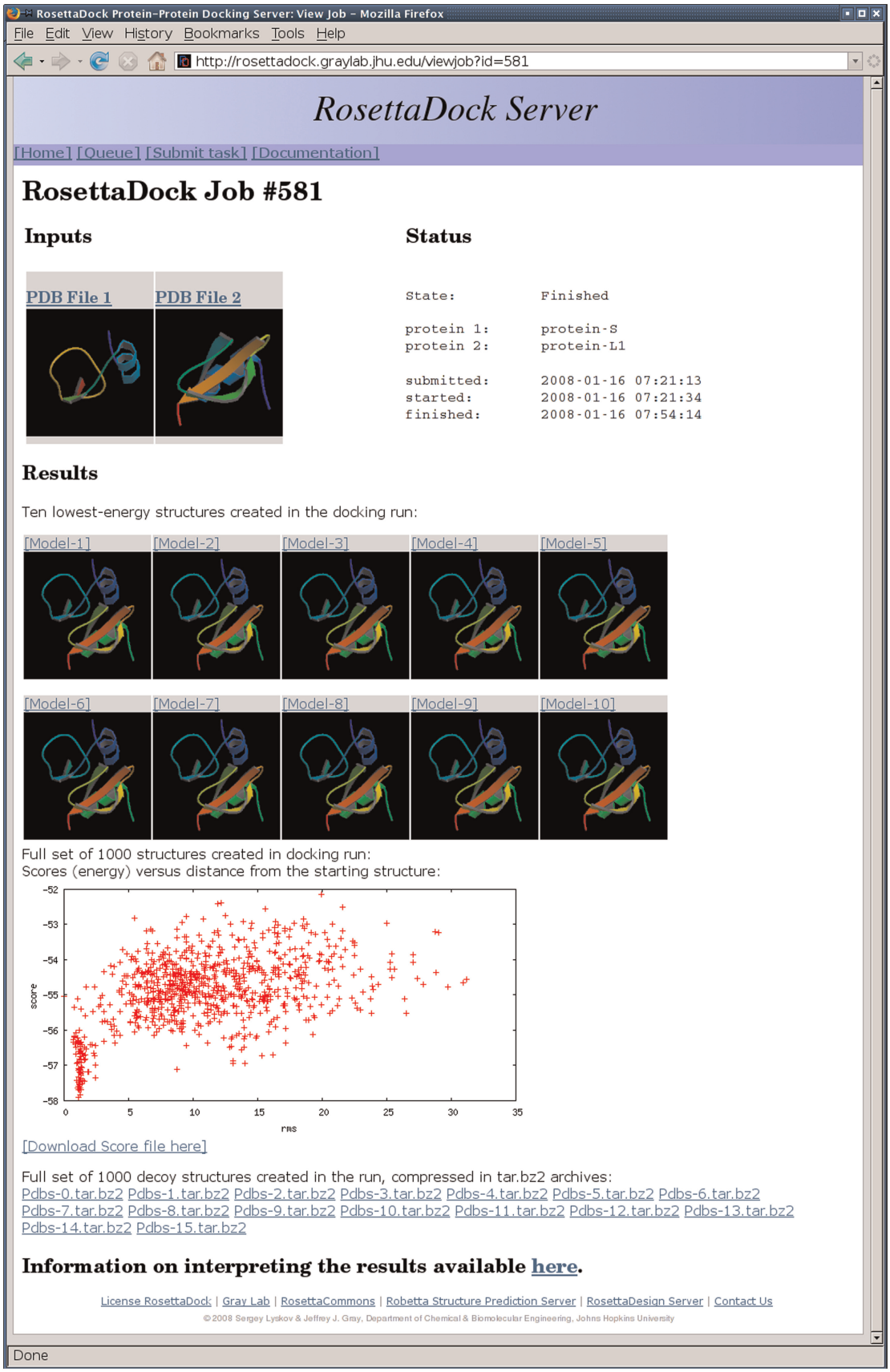

Figure 1. Sample results page. In this example, all 10 low-energy conformations are similar and the score versus r.m.s.d. plot exhibits a binding funnel at ~1 Å from the starting input conformation.

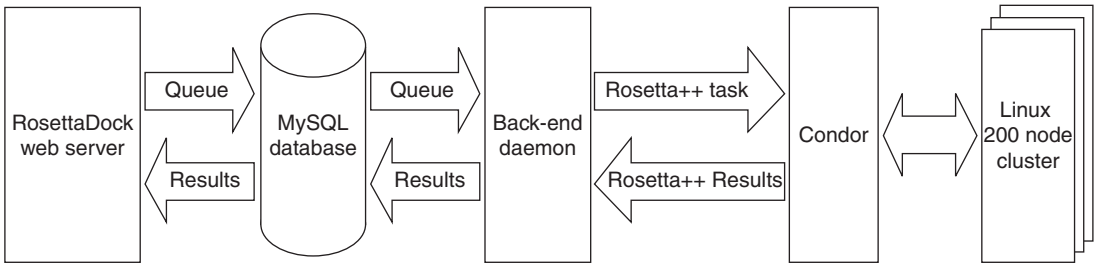

Figure 2. System architecture.

complex structure, a near-native structure was successfully identified in over 60% of cases with low-energy conformations within 10 Å of the lowest-r.m.s.d. superposition of the unbound components onto the bound complex; and in 80% of cases, one of the five top-scoring models correctly captured at least 25% of native residue–residue contacts across the binding interface. The server also incorporates the side-chain refinement techniques of Wang *et al*. (9), which improved the recovery of correct rotameric side-chain conformations and discrimination of near-native complex structures (as measured by *z*-scores).

RosettaDock has also been repeatedly and successfully tested in the CAPRI blind challenge on diverse targets including antigen–antibody pairs, enzyme–inhibitor pairs, regulatory proteins and others (17). In Rounds 3–5, RosettaDock correctly predicted all five targets under 450 total residues to medium or high accuracy, including prediction of the complex of dockerin and cohesion where the dockerin structure was obtained by homology modeling (18). In the recent Rounds 6–12 of CAPRI, two of five targets were predicted correctly using techniques available on the web server [one in combination with the RosettaInterface mode available on Robetta (19)]; other targets pushed the boundaries of RosettaDock's applicability in backbone flexibility and serve as precautions for server users to carefully choose their targets and interpret server results with caution (20). Several of our predictions have been atomically accurate at the interface including many side-chain conformations. For example, the TolB/Pal model (Target 26, an unbound–unbound target) included 47% of the native residue–residue contacts and 1.24 Å interface residue C_α r.m.s.d. to the complex structure. Similarly, the complex of Orc1 with Sir1 was predicted with 46% of the native residue–residue contacts and 1.92 Å interface r.m.s.d. Comparable accuracies were achieved by Shueler-Furman *et al*. (21) and Wang *et al*. (22) using the RosettaDock method with several extensions. Note that the most accurate structure was one of the 10 top-scoring structures, but not necessarily the top-ranked model. Finally, in the recent rounds of CAPRI, Rosetta has been additionally validated by other CAPRI participants. Because of the ability of RosettaDock to refine a local docked conformation to find high-resolution binding modes, other CAPRI participants successfully used RosettaDock for refinement and ranking in the CAPRI experiment. One group produced correct models for the scoring experiment with 30–55% native residue–residue contacts and interface r.m.s.d. ranging from 1.1 to 2.4 Å (23), and another used RosettaDock both before and after additional refinement with steered molecular dynamics (24).

Potential uses of the RosettaDock server

Given biochemical information, such as (but not limited to) mutagenesis data, users can employ software like Pymol (http://pymol.org) to manually orient the two partners in a manner that agrees with the experimental information, and then users can refine the structure using RosettaDock to produce high-resolution structural models of the complex (as in TolB/Pal and Orc1/Sir1 CAPRI targets, which both relied on local docking and biochemical information). These results can then be analyzed with RosettaInterface (25) to test whether the mutagenesis data is recapitulated by the structural docking model and to suggest further mutations for validation. Alternatively, if a reasonable guess of the docking orientation can be determined from homologous structures or complexes in the PDB, that structure can also be refined locally with some accuracy.

If structures of the individual components are not readily available, they can be modeled *de novo* or by homology by using a tool such as the Robetta server (http://robetta.org) (19). We must caution that docking has not been extensively tested with homology structures and it is likely that accumulated errors will frustrate high-resolution predictions. Thus, the incorporation of experimental biochemical information becomes even more important.

We have followed this strategy to use RosettaDock to predict antibody–antigen structures of therapeutic interest to provide hypotheses on a drug mechanism (26) and insights into affinity maturation (27) for complexes, where experimental structures were not available and crystallization presented challenges. RosettaDock has also been used on a family of rotavirus-specific antibodies and the evolution of the neutralizing antibodies was exploited to help validate the models (28). Other examples of RosettaDock application targets range from calcium channels (29) and malaria proteins (30) to antibody Fc interactions (31). The RosettaDock method has been combined with mass spectroscopy (32), cross-linking (32), electron microscopy (33) and homology modeling (30,33–35).

In additional to stand-alone use, RosettaDock can be combined with other docking servers, using its capability

of local searches to refine proposed docking positions. A recent work combines ZDOCK with RosettaDock and re-ranks RosettaDock models with significant success (36). In principle, RosettaDock could be used to refine candidate solutions from any global docking method.

FUTURE DIRECTIONS

Recent work on protein–protein docking has included tailored backbone flexibility (20,22). We hope to expand the server to this type of task (which would require more sophisticated input schemes), however, like global docking, providing this service is limited by the amount of computing resources we are able to donate to the public. More recent work with docking protein ensembles (37) is efficient and we plan to provide backbone flexibility on the server via this technique. Importantly, ensemble docking allows the use of NMR solution-state structures as inputs. Finally, due to inconsistencies in PDB coordinate files, jobs sometimes are unable to complete the RosettaDock program. As issues appear, we are continually implementing various input file validity checks (such as the existing missing backbone atom and protein size checks) toward the goal of clearly reporting to the user all potential errors for immediate correction.

ACKNOWLEDGEMENTS

This study was funded by the National Institutes of Health (R01-GM073151, R01-GM078221). Michael Daily provided some of the documentation for the server and Sidhartha Chaudhury assisted in testing docking runs, inspecting resulting output and reviewing the article. Funding to pay the Open Access publication charges for this article was provided by the National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

1. Gray,J.J. (2006) High-resolution protein-protein docking. *Curr. Opin. Struct. Biol.*, **16**, 183–193.
2. Comeau,S.R., Gatchell,D.W., Vajda,S. and Camacho,C.J. (2004) ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, **20**, 45–50.
3. Tovchigrechko,A. and Vakser,I.A. (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.*, **34**, W310–W314.
4. Chen,R., Li,L. and Weng,Z. (2003) ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, **52**, 80–87.
5. Schneidman-Duhovny,D., Inbar,Y., Nussinov,R. and Wolfson,H.J. (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.*, **33**, W363–W367.
6. Ritchie,D.W. and Kemp,G.J. (2000) Protein docking using spherical polar Fourier correlations. *Proteins*, **39**, 178–194.
7. Gray,J.J., Moughon,S., Wang,C., Schueler-Furman,O., Kuhlman,B., Rohl,C.A. and Baker,D. (2003) Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.*, **331**, 281–299.
8. Lensink,M.F., Mendez,R. and Wodak,S.J. (2007) Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins*, **69**, 704–718.
9. Wang,C., Schueler-Furman,O. and Baker,D. (2005) Improved side-chain modeling for protein-protein docking. *Protein Sci.*, **14**, 1328–1339.
10. Gray,J.J., Moughon,S.E., Kortemme,T., Schueler-Furman,O., Misura,K.M., Morozov,A.V. and Baker,D. (2003) Protein-protein docking predictions for the CAPRI experiment. *Proteins*, **52**, 118–122.
11. Kortemme,T., Morozov,A.V. and Baker,D. (2003) An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.*, **326**, 1239–1259.
12. Lazaridis,T. and Karplus,M. (2000) Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.*, **10**, 139–145.
13. Dunbrack,R.L. Jr. and Cohen,F.E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.*, **6**, 1661–1681.
14. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
15. Liu,Y. and Kuhlman,B. (2006) RosettaDesign server for protein design. *Nucleic Acids Res.*, **34**, W235–W238.
16. London,N. and Schueler-Furman,O. (2007) Assessing the energy landscape of CAPRI targets by FunHunt. *Proteins*, **69**, 809–815.
17. Janin,J., Henrick,K., Moult,J., Eyck,L.T., Sternberg,M.J., Vajda,S., Vakser,I. and Wodak,S.J. (2003) CAPRI: a critical assessment of predicted interactions. *Proteins*, **52**, 2–9.
18. Daily,M.D., Masica,D., Sivasubramanian,A., Somarouthu,S. and Gray,J.J. (2005) CAPRI rounds 3–5 reveal promising successes and future challenges for RosettaDock. *Proteins*, **60**, 181–186.
19. Kim,D.E., Chivian,D. and Baker,D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.*, **32**, W526–W531.
20. Chaudhury,S., Sircar,A., Sivasubramanian,A., Berrondo,M. and Gray,J.J. (2007) Incorporating biochemical information and backbone flexibility in RosettaDock for CAPRI rounds 6–12. *Proteins*, **69**, 793–800.
21. Schueler-Furman,O., Wang,C. and Baker,D. (2005) Progress in protein-protein docking: atomic resolution predictions in the CAPRI experiment using RosettaDock with an improved treatment of side-chain flexibility. *Proteins*, **60**, 187–194.
22. Wang,C., Schueler-Furman,O., Andre,I., London,N., Fleishman,S.J., Bradley,P., Qian,B. and Baker,D. (2007) RosettaDock in CAPRI rounds 6–12. *Proteins*, **69**, 758–763.
23. Wiehe,K., Pierce,B., Tong,W.W., Hwang,H., Mintseris,J. and Weng,Z. (2007) The performance of ZDOCK and ZRANK in rounds 6–11 of CAPRI. *Proteins*, **69**, 719–725.
24. Heifetz,A., Pal,S. and Smith,G.R. (2007) Protein–protein docking: progress in CAPRI rounds 6–12 using a combination of methods: the introduction of steered solvated molecular dynamics. *Proteins*, **69**, 816–822.
25. Kortemme,T., Kim,D.E. and Baker,D. (2004) Computational alanine scanning of protein-protein interfaces. *Sci. STKE*, 2004, pl2
26. Sivasubramanian,A., Chao,G., Pressler,H.M., Wittrup,K.D. and Gray,J.J. (2006) Structural model of the mAb 806-EGFR complex using computational docking followed by computational and experimental mutagenesis. *Structure*, **14**, 401–414.
27. Sivasubramanian,A., Maynard,J.A. and Gray,J.J. (2008) Modeling the structure of mAb 14B7 bound to the anthrax protective antigen. *Proteins*, **70**, 218–230.
28. McKinney,B.A., Kallewaard,N.L., Crowe,J.E. Jr. and Meiler,J. (2007) Using the natural evolution of a rotavirus-specific human monoclonal antibody to predict the complex topography of a viral antigenic site. *Immunome Res.*, **3**, 8.
29. Hulme,J.T., Yarov-Yarovoy,V., Lin,T.W.C., Scheuer,T. and Catterall,W.A. (2006) Autoinhibitory control of the CaV1.2 channel by its proteolytically processed distal C-terminal domain. *J. Physiol.*, **576**, 87–102.
30. Bertonati,C. and Tramontano,A. (2007) A model of the complex between the PfEMP1 malaria protein and the human ICAM-1 receptor. *Proteins Struct. Func. Genet.*, **69**, 215–222.
31. Sprague,E.R., Wang,C., Baker,D. and Bjorkman,P.J. (2006) Crystal structure of the HSV-1 Fc receptor bound to Fc reveals a mechanism for antibody bipolar bridging. *PLoS Biol.*, **4**, 0975–0986.
32. Schulz,D.M., Kalkhof,S., Schmidt,A., Ihling,C., Stingl,C., Mechtler,K., Zschoernig,O. and Sinz,A. (2007) Annexin A2/P11 interaction: new insights into annexin A2 tetramer structure by chemical crosslinking, high-resolution mass spectrometry, and computational modeling. *Proteins Struct. Func. Genet.*, **69**, 254–269.

33. Diemand,A.V. and Lupas,A.N. (2006) Modeling AAA+ ring complexes from monomeric structures. *J. Struct. Biol.*, **156**, 230–243.
34. Cestèle,S., Yarov-Yarovoy,V., Qu,Y., Sampieri,F., Scheuer,T. and Catterall,W.A. (2006) Structure and function of the voltage sensor of sodium channels probed by a β-scorpion toxin. *J. Biol. Chem.*, **281**, 21332–21344.
35. Lacy,D.B., Lin,H.C., Melnyk,R.A., Schueler-Furman,O., Reither,L., Cunningham,K., Baker,D. and Collier,J.R. (2005) A model of anthrax toxin lethal factor bound to protective antigen. *Proc. Natl Acad. Sci. USA*, **102**, 16409–16414.
36. Pierce,B. and Weng,Z. (2008) A combination of rescoring and refinement significantly improves protein docking performance. *Proteins*, available in EarlyView; doi:101002/prot.21920.
37. Chaudhury,S. and Gray,J.J. (2008) Conformer selection and induced fit in flexible backbone protein-protein docking using computational and NMR ensembles. Manuscript in review. *J. Mol. Biol.*